AD-A039 182 CALIFORNIA UNIV LOS ANGELES DEPT OF SYSTEM SCIENCE F/G 12/1
PARAMETER ESTIMATION IN STOCHASTIC DIFFERENTIAL SYSTEMS: THEORY--ETC(U)
JAN 77 A V BALAKRISHNAN AF-AFOSR-2492-73

UNCLASSIFIED AFOSR-TR-77-0430 NL

| OF |
ADA039182

END

DATE
FILMED
5 - 77

AFOSR - TR - 77 - 0430

AD A039182

Parameter Estimation in Stochastic Differential Systems: Theory and Application

A. V. Balakrishnan

To be Published in
Advances in Statistics 5ed.
Edited by P. R. Krishnaiah
Academic Press 1977

DDC
MAY 9 1977
C

AD No.
DDC FILE COPY

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19 REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 18 AFOSR - TR - 77 - 0430	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER 9
4. TITLE (and Subtitle) 6 PARAMETER ESTIMATION IN STOCHASTIC DIFFERENTIAL SYSTEMS: THEORY AND APPLICATION		5. TYPE OF REPORT & PERIOD COVERED Interim rept,
		6. PERFORMING ORG. REPORT NUMBER [redacted]
7. AUTHOR(s) 10 A. V. Balakrishnan		8. CONTRACT OR GRANT NUMBER(s) 15 AF- AFOSR ~~73~~-2492-73
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of California, Los Angeles System Science Department Los Angeles, CA 90024		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 17 16 2304/A1
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientifc Research (NM) Building 410 Bolling AFB Washington, D.C. 20332		12. REPORT DATE 11 January 1977
		13. NUMBER OF PAGES 51
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 12 53 p.		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

To be Published in Advances in Statistics 5ed.
Edited by P. R. Krishnaiah
Academic Press 1977

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

This paper

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

~~We~~ present a theory of estimation of parameters in linear stochastic differential equations based on time-continuous observation. It ~~We~~ uses a white noise model to represent observation errors (in contrast to a Wiener process model). The application is to the problem of identifying aircraft as well as turbulence (wind-gust) parameters from flight test data. Results obtained on actual data ~~(not simulated data)~~ are presented.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE UNCLASSIFIED 405 213

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

BEST AVAILABLE COPY

Parameter Estimation in Stochastic Differential Systems: Theory and Application

A. V. Balakrishnan †

Abstract We present a theory of estimation of parameters in linear stochastic differential equations based on time-continuous observation. We use a white noise model to represent observation errors (in contrast to a Wiener process model). The application is to the problem of identifying aircraft as well as turbulence (wind-gust) parameters from flight test data. Results obtained on actual data (not simulated data) are presented.

† Research supported in part under Grant No. 73-2492, Applied Mathematics Divn., AFOSR, USAF.

1. Introduction

The estimation problem in essence is the following. We have an observed process $y(t)$ ($n \times 1$ matrix function) which has the form

$$y(t) = S(\theta, t) + N(t) \qquad 0 < t < T \tag{1.1}$$

where 'θ' denotes a vector of unknown parameters which we want to estimate, $S(\theta, t)$ being a stochastic process ("signal") which is completely specified once θ is specified (by means of a stochastic differential system, for example) and $N(t)$ is a stochastic process which models the errors (that remain even after all 'systematic' errors, such as bias and calibration errors, have been accounted for). There is much evidence to suggest that the noise process may be well modelled as Gaussian, and independent of the signal process. This is a basic assumption thruout in this paper.

Under the title of "System Identification" there is a large engineering literature dealing with such problems. This is well documented in the proceedings of three symposia [1] devoted exclusively thereto. In the bulk of this literature, the process $S(\theta, t)$ is taken to be deterministic, in which case the estimation is largely treated as a 'Least Squares' problem of minimising

$$\int_0^T ||y(t) - S(\theta, t)||^2 dt$$

over a predetermined admissible set of parameters θ. Where the stochastic signal case is considered, it is reduced to the time-discrete version of (1.1):

$$y_n = S_n(\theta) + N_n \tag{1.2}$$

for the reason that the continuous time is mathematically too difficult to handle, and anyhow, in digital computer processing (as is the rule), it is so discretized in the A-D conversion process anyway. This is indeed true; but the authors invariably proceed to make the assumption that the noise samples $\{N_n\}$ are mutually independent. But this requires that the sampling rate (in the periodic sampling of the data) be not more than twice the postulated 'bandwidth' of the noise, itself actually unknown. Indeed in most practical cases the sampling rate is far higher than twice the bandwidth. To meet this objection, one may then allow the $\{N_n\}$ to be correlated. But then the correlation function must be known, and anyone with experience in handling real data can easily appreciate that it is unrealistic to require that much knowledge of the noise process, even if the complication in the theory can be borne.

We maintain, in any event, that it is much better to work with the time-continuous model (1.1), allowing as high a sampling rate in the processing as the A-D converter is designed for. But in the time-continuous model we are faced with another problem. The basic tool in estimation is the likelihood functional (for fixed parameters) which is based on the Radon-Nikodym derivative of the probability measure induced by the process by $y(\cdot)$ to that induced by the noise process $N(t)$. But this derivative is too difficult to calculate even when the precise spectrum of $N(\cdot)$ is known, which it is not. What we can assert for sure is that the bandwidth of noise $N(t)$ is much larger than that of the process $S(\theta; t)$, which is essential in order that the measuring instrument does not 'distort' the signal. At this point it was customary in the earlier engineering literature to introduce "white noise" in a formal way as a stationary stochastic process with constant spectral density to represent the 'large bandwidth' nature of N(t). With the advances in the theory of diffusion processes using the Ito integral, it became fashionable to use a Wiener process model as being "more rigorous" [2]. Thus we replace (1.1) by

$$Y(t) = \int_0^t S(\theta,\sigma)d\sigma + W(t) \tag{1.3}$$

where W(t) is a Wiener process. We can then exploit the well-developed machinery of martingales and Ito intergrals. In fact the likelihood function can then be expressed as: (see [2]):

$$\text{Exp} - 1/2\left\{ \int_0^T ||\hat{S}(\theta,t)||^2 dt - 2\int_0^T [\hat{S}(\theta,t), dY(t)]\right\} \tag{1.4}$$

where $\hat{S}(\theta,t)$ is the best mean square estimate of $S(\theta,t)$ given the sigma-algebra generated by $Y(s)$, $s \leq t$. This formula can be justly considered as one of the triumphs of the Ito theory, the key to the success being the appearance of the Ito-integral in the second form of (1.4). This integral is defined on the basis that Y(t) is of unbounded variation with probability one. Of course no physical instrument can produce such a waveform. To calculate it, given the actual observation (1.1), we can "retrace" our steps back from (1.3) and use

$$y(t)dt$$

in place of dY(t). But this is totally incorrect, unless $S(\theta,t)$ is deterministic, and any minimisation procedure based on it leads to erroneous results. This point is not appreciated by authors using (1.3) as "more rigorous", perhaps because they have not had occasion to actually calculate anything based on real data. In any data generated by digital computer simulation, which must perforce employ the discrete version (1.2), this point can be completely masked and hence never appreciated.

Faced with this difficulty we have to examine more precisely the model again, to see a physically more meaningful way of exploiting the fact that the noise bandwidth is large compared to the signal bandwidth. What is needed is the 'asymptotic form' of the likelihood functional as the bandwidth expands to infinity in an arbitrary manner.

Such a theory has been developed by the author using a precise notion of white noise. This is explained in Section 2. Based on this theory we derive a likelihood functional in Section 3. It turns out that formula (1.4) is replaced by

$$\operatorname{Exp} - 1/2 \left\{ \int_0^T \|\hat{S}(\theta,t)\|^2 dt - 2 \int_0^T \hat{S}(\theta,t)\, y(t)dt + \int_0^T \left(\widehat{\|S(\theta,t)\|^2} - \|\hat{S}(\theta,t)\|^2\right)dt \right\} \tag{1.5}$$

where $\wedge$ denotes conditional expectation given the data upto time t. Note that a third term appears which can also be expressed as:

$$\int_0^T \|S(\theta,t) - \hat{S}(\theta,t)\|^2 dt$$

and in the case where $S(\theta,t)$ is Gaussian, this reduces to

$$\int_0^T E[\|S(\theta,t) - \hat{S}(\theta,t)\|^2]dt$$

being thus the integral of the mean square error in estimation of the signal $S(\theta,t)$ from the observation upto time t. When the signal process can be described in terms of stochastic differential equations, whether finite or infinite dimensional advantage can be taken of the fact that the mean square error can be evaluated by solving a Riccati equation. Section 4 is devoted to this specialization. Section 5 deals with the application to the problem of stability and control derivatives from flight test data taking turbulence into account. The algorithms used and results obtained on actual flight data are included.

2. WHITE NOISE: BASIC NOTIONS

Let H denote a real separable Hilbert Space and let

$$W = L_2\,[0,\ T;\ H],\quad 0 < T < \infty,$$

denote the real Hilbert Space of H-valued weakly measurable functions $u(\cdot)$ such that

$$\int_0^T [u(t),\ u(t)]dt < \infty$$

with inner-product defined by

$$[u,\ v] = \int_0^T [u(t),\ v(t)]dt$$

Let μ_G denote Gauss measure on W (on the cylinder sets with finite dimensional Borel basis) with characteristic function

$$C_G(h) = \text{Exp} - 1/2[h,\ h],\ h \in W.$$

Elements of W under this (finitely additive) measure will be 'white noise sample functions', denoted ω. This terminology appears to have the sanction of usage; see Skorokhod [3] for example. It is essential for us that W is an L_2-space over a finite interval.

Any function $f(\cdot)$ defined on W into another Hilbert Space H_r such that the inverse images of Borel sets in H_r are cylinder sets will be called a 'tame' function. See Gross [4]. As is readily seen, the class of tame functions is a linear class. Since the inverse image of the whole space H_r

BEST AVAILABLE COPY

must be cylindrical, it is clear that any tame function has the form $f(P\omega)$ where P is a finite dimensional projection.

To introduce the notion of a 'random variable' let us first confine ourselves to the case where H_r is finite dimensional: $H_r = R^n$ say. We introduce a metric into the linear space of tame functions by:

$$|||f-g||| = \int_W \frac{||f-g||}{1 + ||f-g||} \, d\mu_G$$

and then complete the space, the completion yielding a Frechet Space. Every element of the completed space is called a 'random variable' and if ζ denotes such an element and $f_n(\omega)$ a corresponding Cauchy sequence in probability, then we define the corresponding 'distribution function' or probability measure on R^n to be that induced by the characteristic function

$$C_\zeta(h) = \underset{n}{\text{limit}}\ E(e^{i[f_n(\omega),\, h]}) \tag{2.0}$$

The latter limit exists (uniformly on bounded sets of $R^n = H_r$).

In the case where H_r is no longer finite dimensional, we shall still identify Cauchy sequences in probability of tame functions as "weak random variables". The limit in (2.0) still holds, uniformly on bounded sets in H_r, but the limit may in general only define a "weak distribution" on H_r. We recall in this connection the Sazonov theorem [5] that the limit is the characteristic function of a probability measure if and only if it is continuous in the trace-norm topology ('S-topology' see below). This is automatically the case if the sequence is Cauchy in the mean square sense; and we shall then drop the qualification "weak".

Let $f(\omega)$ be any Borel measurable function mapping W into H_r. Then $f(P\omega)$ is tame for every finite-dimensional projection operator P. Let $\{P_n\}$ denote a sequence of finite dimensional projections converging strongly to the Identity; the sequence may be assumed to be monotone. If the sequence $f(P_n\omega)$ is Cauchy in probability, then we may associate a (weak, ingeneral) random variable with $f(\cdot)$. Let us denote it by $\tilde{f}$ (a notation used by Gross). This limit of course can depend on the particular projection sequence chosen. Of primary interest to us are those functions $f(\cdot)$ for which $\{f(P_n\omega)\}$ is Cauchy in probability for every such sequence of finite dimensional projections and moreover such that all such Cauchy sequences are equivalent so that the limit random variable $\tilde{f}$ is unique. In that case we say that $f(\omega)$ is a (weak) random variable. We shall use the term "random variable" if the corresponding measure is countably additive; we shall be dealing in the sequel only with mean square convergence, when this will be automatic.

The simplest function one can consider is perhaps the linear function:

$$f(\omega) = L\omega$$

where L is a linear bounded transformation mapping W into H_r, where we now allow H_r to be infinite dimensional. Then it is easy to see that if L is Hilbert-Schmidt, then $\{LP_m\omega\}$ is Cauchy in the mean square sense, and $L\omega$ is a random variable. Convsersely L must be H.S. if $L\omega$ is to be a random variable.

What is the class of functions which are random variables? To answer this question, at least in part, let us introduce the S-topology on W: this is the (locally convex) topology induced by seminorms of the form:

$$\rho(\omega) = \sqrt{[S\omega,\omega]} \tag{2.1}$$

where here (and hereinafter) S will denote a self-adjoint, non-negative definite trace-class operator on W into W. For the case where $H_r = R^1$, Gross [4] has given a sufficient condition: $f(\cdot)$ is a random variable if it is uniformly continuous in the S-topology. Uniform continuity means that given $\varepsilon > 0$, we can find $\rho(\cdot)$ such that

$$||f(x) - f(y)|| < \varepsilon \quad \text{for all } x, y \text{ such that } \rho(x-y) < 1.$$

Unfortunately Gross does not seem to discuss non-trivial examples of functions satisfying this condition. Here we shall give a sufficient condition for a class of random variables with finite second moment.

Theorem 2.1

Let $p_n(\omega)$ denote a homogeneous polynomial of degree n mapping W into H_r. Suppose it is continuous at the origin in the S-topology. Let P denote any finite dimensional projection.

$$\sup_P E(||p_n(P\omega)||^2) < \infty \tag{2.2}$$

where the supremum is taken over the class of all finite dimensional projections. Conversely, if (2.2) holds, then $p_n(\cdot)$ is continuous at the origin in the S-topology.

Proof We begin with a simple but useful Lemma.

Lemma 2.1

Suppose $p_n(\cdot)$ is continuous in the S-topology at the origin. Then there exists a seminorm in the S-topology:

$$\rho(\omega) = \sqrt{[S\omega,\omega]} \tag{2.3}$$

such that

$$||p_n(\omega)|| \leq M\, \rho(\omega)^n \tag{2.4}$$

where M is a constant. Conversely if (2.4) holds, then $p_n(\omega)$ is continuous in the S-topology at the origin.

Proof

Continuity in the S-topology at zero implies this: given $\varepsilon > 0$ we can find a seminorm of the form (2.3) such that

$$||p_n(\omega)|| < \varepsilon \quad \text{for all } \omega \text{ such that } \rho(\omega) \leq \delta \tag{2.5}$$

Hence for any ω for which

$$\rho(\omega) \neq 0,$$

we have that

$$||p_n(\frac{\delta\omega}{\rho(\omega)})|| < \varepsilon$$

or by the homogeneity of $p_n(\cdot)$,

$$||p_n(\omega)|| < (\frac{\varepsilon}{\delta^n})\ \rho(\omega)^n,\ \rho(\omega) \neq 0$$

If $\rho(\omega) = 0$, then for any positive number k,

$$\rho(k\omega) = 0$$

and hence from (2.5)

$$||p_n(\omega)|| < \varepsilon|k^n$$

for all $k > 0$ and hence

$$p_n(\omega) = 0$$

Hence 2.4 holds. The converse is obvious.

Proof of Theorem

Corresponding to a finite dimensional projections P, we can find an orthonormal basis $\{\phi_i\}$ such that P is the projection operator corresponding to the space spanned by ϕ_i, $i = 1,2,\ldots m$. Let

$$p_n(\omega) = k_n(\omega,\ldots\omega)$$

$k_n(\ldots)$ being the symmetric n-linear form, corresponding to $p_n(\cdot)$.

Then

$$p_n(P\omega) = \sum_{i_1=1}^{m} \cdots \sum_{i_n=1}^{m} a_{i_1,\ldots i_n} \zeta_{i_1}\ldots\zeta_{i_n} \tag{2.6}$$

where

$$a_{i_1,\ldots i_n} = k_n(\phi_{i_1},\ldots\phi_{i_n})$$

$$\zeta_i = [\phi_i,\omega].$$

The $\{\zeta_i\}$ is a sequence of independent zero-mean unit variance Gaussians and (2.6) defines a tame function. Moreover we can readily calculate (by expressing (2.6) in terms of Hermite polynomial for instance) that

$$E(||p_n(P\omega)||^2)$$

$$= \sum_{\nu=0}^{[n/2]} \left(\frac{n!}{(n-2\nu)!2^\nu \nu!}\right)^2 \sum_{i_{2\nu+1}=1}^{m} \sum_{i_n=1}^{m} || \sum_{i_1=1}^{m} \cdots \sum_{i_\nu=1}^{m}$$

$$a_{i_1,i_1,\ldots i_\nu,i_\nu,i_{2\nu+1},\ldots i_n}||^2 \tag{2.7}$$

But from Lemma 2.1, we have that

$$||p_n(P\omega)||^2 \leq [S_m\omega,\omega]^n \tag{2.8}$$

where

$$S_m = P\ SP$$

and is of course trace-class and finite dimensional.

Hence

$$E[||p_n(P\omega)||^2] \leq E([S_m\omega,\omega]^n) \tag{2.9}$$

Let ψ_k, $k = 1,\ldots\nu$, be the orthonormalized eigen-vectors of S_m with corresponding non-zero eigen-values λ_k.

Then

$$[S_m\omega,\omega] = \sum_1^\nu \lambda_i[\psi_i,\omega]^2$$

and we have

$$E([S_m\omega,\omega]^n) = f(\mathrm{Tr}.S_m,\ \mathrm{Tr}.S_m^2,\ldots\mathrm{Tr}.S_m^n)$$

where $f(\cdots)$ is a fixed continuous function. Of course

$$\mathrm{Tr.}\ S_m^j$$

is monotone in m for each j and converge to

$$\mathrm{Tr.}\ S^j$$

Hence it follows that

$$E[||p_n(P\omega)||^2] < \infty$$

for all finite dimensional projections.

To prove the converse, suppose (2.2) holds. The (2.7) holds for every m, and taking $\nu = 0$ therein, we obtain that

$$\sum_{i_1}^{\infty} \cdots \sum_{i_r}^{\infty} ||k_n(\phi_{i_1}, ..\phi_{i_n})||^2 < \infty \tag{2.10}$$

for every orthonormal sequence $\{\phi_i\}$. Hence $p_n(\cdot)$ is Hilbert-Schmidt. Of course

$$||p_n(\omega)||^2 \leq M\ ||\omega||^{2n} \tag{2.11}$$

Define now S by:

$$[S\omega,\omega] = (||p_n(\omega)||^2)^{1/n}$$

Then S is Hilbert-Schmidt by (2.10).

For any finite dimensional projection P,

$$\begin{aligned} E[SP\omega,P\omega] &= E[PSP\omega,\omega] \\ &= E((||p_n(P\omega)||^2)^{1/n}) \end{aligned}$$

and hence

$$\sup_{P} E[PSP\omega,\omega] < \infty$$

But taking the orthonormal basis of eigen-vectors of S, it follows that S is trace-class.

It follows from Theorem 2.1 that if a homogeneous polynomial is uniformly continuous in the S-topology, the corresponding random variable has finite second moment.

For a homogeneous polynomial of degree 2 with range in R^1($H_r = R^1$ or R^m more generally) we can prove that continuity at the origin in the S-topology is sufficient to make it a random variable. For from (2.7) we have

$$E[||p_2(P\omega)||^2] = \sum_1^m \sum_1^m |k_2(\phi_i,\phi_j)|^2 + |\sum_1^m k_2(\phi_i,\phi_i)|^2 < \infty$$

and hence

$$\sum_1^\infty |k_2(\phi_i,\phi_i)| < \infty$$

for any orthonormal system. Hence it follows that

$$E[||p_2(P_n\omega) - p_2(P_m\omega)||^2]$$

is Cauchy. This suffices for our purposes here. See [6] for more, and in particular the relation to multiple Ito integrals.

3. RADON-NIKODYM DERIVATIVES OF WEAK DISTRIBUTIONS.

Let ω denote white noise samples as in Section 2 and let

$$y(\omega) = f(\omega) + \omega \tag{3.1}$$

where $f(\cdot)$ is a random variable mapping into W; then $\{y(P_m \omega)\}$ is a Cauchy sequence in probability (being the sum of two such sequences) and the limit is independent of the particular sequence $\{P_m\}$ chosen. Hence $y(\omega)$ induces a weak distribution on W. Call it μ_y. As finitely additive measures, μ_y is said to be absolutely continuous with respect to μ_G if given any $\varepsilon > 0$, we can find $\delta > 0$ such that for any cylinder set C,

$$\mu_y(C) < \varepsilon$$

as soon as

$$\mu_G(C) < \delta.$$

The definition of the derivative however is more involved. For our purposes, we shall be concerned with the case where the derivative is a random-variable. That is to say, there exists a function $f(\omega)$ mapping W into R^1 such that $f(\omega)$ is a random variable and for any cylinder set C:

$$\mu_y(C) = \lim_m \int_C f(P_m \omega) d\mu_G$$

where $\{P_m\}$ is any monotone sequence of finite dimensional projections converging strongly to the identity.

Let

$$W_s = L_2 \, [(0,T); H_s]$$

where H_s is a separable Hilbert space. Let μ^s_G denote Gauss measure thereon, and let ω_s denote points in W_s. (The subscript s stands for 'signal'). Let

$$W_2 = W_s \otimes W$$

the Cartesian product and induce the product Gauss measure μ_2 on W_2:

$$\mu_2(C_s \otimes C) = \mu^s_G(C_s)\ \mu_G(C)$$

where C_s is a cylinder set in W_s and C a cylinder set in W. Define Denote points in W_2 by ω_2:

$$\omega_2 = \begin{bmatrix} \omega_s \\ \omega \end{bmatrix}.$$

Let

$$y(\omega_2) = f(\omega_s) + \omega \tag{3.2}$$

where $f(\cdot)$ is a random variable mapping W_s into W. Let μ_y denote again the (finitely additive) measure induced by $y(\cdot)$. We wish to prove the absolute continuity of the measure $\mu_y(\cdot)$ with respect to the measure $\mu_G(\cdot)$ and to find the corresponding derivative.

For the Wiener process version of (1.2), such a result appears to have been first developed by Duncan [7] for the case where $f(\omega_s)$ is a diffusion process. See also [8] as may be expected, our result has a superficial similarity to Stratanovich version [9, eq. 12].

Let H be finite dimensional: $H = R^n$.

Theorem 2.1 Let $f(\omega_s)$ denote a random variable mapping W_s into W such that

$$E(||f(\omega_s)||^2) < \infty \tag{3.3}$$

Let $y(\omega_2)$ be defined by (3.2).

Then μ_y is absolutely continuous with respect to μ_G and the derivative is a random variable (white noise integral), corresponding to the function $g(\omega)$ defined by:

$$g(\omega) = \int_W (\text{Exp} - 1/2\{||x||^2 - 2\,[x,\omega]\})\, d\mu_S \tag{3.4}$$

where x is a dummy variable denoting points in W, and $\mu_s(\cdot)$ is the countably additive measure induced by $f(\cdot)$ on the Borel sets of W. More precisely:

$$\lim_m \; E(\text{Exp}\; i\,[f(P_m\,\omega_s),h])$$

$$= C(h)$$

$$= \int_W e^{i[\omega,h]} d\mu_s$$

where P_m is any monotone sequence of finite dimensional projections converging strongly to the identity.

Proof

With μ_s denoting the (countably additive) measure induced by $f(\omega_s)$ on W, define for each ω:

$$g(\omega) = \int_W \text{Exp} - 1/2\{||x||^2 - 2\,[x,\omega]\}\, d\mu_S$$

This is well defined since the integrand is continuous in x, non-negative, and bounded by

$$\operatorname{Exp} 1/2 \; ||\omega||^2.$$

Moreover $g(\omega)$ is actually a continuous functional on W. For, given $\varepsilon > 0$, we can find a closed bounded K_ε such that

$$\mu_S(K_\varepsilon) > 1 - \varepsilon.$$

Then

$$\int_{K_\varepsilon} \operatorname{Exp} - 1/2 \; \{||s||^2 - 2\,[s,\omega]\} \; d\mu_S$$

is continuous in ω and on the complement K_ε, the integral is

$$\leq \; (\exp \frac{||\omega||^2}{2}) \; \varepsilon$$

Now let us show that $g(\omega)$ is a random variable. Let $\{P_m\}$ denote a monotone sequence of finite dimensional projections on W strongly convergent to the identity. Let $\{\phi_i\}$ denote a corresponding orthonormal basis, with the range of P_m being the span of the first m members of the sequence. Let us note that we can write

$$g(P_m\omega) = \int_W (\operatorname{Exp}. - 1/2 \; ||x - P_m x||^2 - 1/2\{||P_m x||^2 - 2\,[P_m x,\omega]\}).\; d\mu_S$$

and hence is

$$\leq \int_W \operatorname{Exp} - 1/2\{||P_m x||^2 - 2\,[P_m x,\omega]\} \; d\mu_S.$$

Let

$$g_m(\omega) = \int_W \text{Exp} - 1/2 \{||P_m x||^2 - 2[P_m x,\omega]\} \, d\mu_S$$

$$= g_m(P_m \omega)$$

Then

$$\int_W g_m(\omega) d\mu_G = \int_W \left(\int_W \text{Exp} - 1/2\{||P_m x||^2 - 2 [P_m x,\omega]\} \, d\mu \right) d\mu_S$$

$$= 1$$

Next

$$E(|g(P_m\omega) - g_m(\omega)|)$$

$$= \int_W \int_W (1 - \text{Exp} - 1/2||x - P_m x||^2) \text{ Exp} - 1/2\{||P_m x||^2 - 2 [P_m x,\omega]\} \, d\mu_S . d\mu_G$$

$$= \int_W (1 - \text{Exp} - 1/2||x - P_m x||^2) d\mu_S \tag{3.5}$$

$$< \varepsilon \qquad \text{for all } m > m(\varepsilon).$$

Hence the convergence properties of $\{g(P_m\omega)\}$ are the same as that of $\{g_m(\omega)\}$. The latter sequence is a martingale. At this point rather than repeat traditional arguments, we shall exploit them and thereby also show the connection to the Wiener process version. Thus let

$$[y(\omega), \phi_i] = y_i = x_i + \zeta_i$$

where

$$x_i = [x, \phi_i]; \quad \zeta_i = [\omega, \phi_i]$$

Here the ζ_i, $i = 1,..n$, for any finite n are independent zero mean, unit variance Gaussians. We can create a "probability space" with a countably additive measure on it such that for any finite number of co-ordinates we have the same distributions: namely R^∞ for the space, and the sigma-algebra β generated by cylinder sets, for the Borel sets. Equivalently, we could use $C[0,T]$ the Banach space of continuous functions with range in R^n, (with the usual sup norm) as the space by defining the mapping W into $C[0,T]$ by:

$$S(t) = \int_0^t x(\sigma)d\sigma \qquad 0 \leq t \leq T$$

and $W(t)$ to be standard Wiener process on $C[0,T]$ and defining

$$Y(t) = \int_0^t x(\sigma)d\sigma + W(t) \tag{3.6}$$

with the Wiener measure and the measure induced by $S(\cdot)$ independent. In this way we get a "co-ordinate free" representation, and we note that the variables

$$\int_0^T [\phi_i(t), dY(t)] = y_i$$

have the same finite dimensional distributions as before. Moreover the variables $g_m(\omega)$ have a corresponding interpretation and have the same distribution for any finite m, and under the condition (3.3), we know that the measure induced by $Y(\cdot)$ is absolutely continuous with respect to Wiener measure, the martingale sequence converging to it in the mean of order one. The derivative itself is given by (see Duncan [7]) by:

$$E_x[\text{Exp} - 1/2 \left[\int_0^T x(t)^2 dt - 2\int_0^T x(t)dW(t)\right] \tag{3.7}$$

where $E_x[\ \ ]$ denotes expectation with respect to the measure induced by the process $S(\cdot)$ on $C[0,T]$; the (Ito) integral (the processes being independent).

$$\int_0^T x(t)dW(t)$$

being the same as:

$$\sum_1^\infty x_i\, \zeta_i \text{ where } x_i = \int_0^T [\phi_i(t),\, dS(t)]$$

$$\zeta_i = \int_0^T [\phi_i(t), dW(t)].$$

We have thus proved that $g(P_m\omega)$ is Cauchy in the mean of order one; and such sequences are equivalent as we change basis. Moreover, it readily follows that for cylinder sets C:

$$\mu_y(C) = \underset{m}{\text{limit}} \int_C g_m(\omega)\, d\mu_G$$

$$= \underset{m}{\text{limit}} \int_C g(P_m\omega) d\mu_G$$

This concludes the proof of the Theorem.

Corollary

For any t, $0 \leq t \leq T$, let

$$\underline{W}(t) = L_2\,[[0,t];\, R^n]$$

$$\underline{W}_S(t) = L_2\,[[0,t];\, H]$$

Let μ_G^t denote the Gauss measure on $\underline{W}(t)$ and μ_S^t similarly the projection of μ_S on the sub-sigma algebra of Borel sets in $\underline{W}(t)$. Then the statement of the theorem applied to measures on $\underline{W}(t)$ reads:

$$g(t;\, P(t)\omega) = \int_W \operatorname{Exp} - 1/2\, \{||P(t)x||^2 - 2\,[P(t)x,\omega]\}\, d\mu_S$$

where P(t) denotes the projection of $\underline{W}$ on $\underline{W}(t)$.

Proof The proof is immediate. We state it rather to note that we cannot take derivatives (with respect to t) in this formula as we can in the Wiener process version.

Remark The Theorem holds for any countably additive measure μ_S on the Borel sets of W, not necessarily generated by a random variable $f(\omega_s)$.

Let us note that the main virtue of the theorem is not so much the formula (3.4) but rather that the derivative is a random variable. The latter has been proved for a related but more general problem in [10] under additional assumptions. We explore this in the next section.

The 'Linear' Case.

Mostly to illustrate the ideas involved, let us consider the special case where $f(\omega_s)$ is linear. Thus let

$$y(\omega_2) = L\,\omega_s + \omega \tag{3.8}$$

where now we allow H in the definition of W to be infinite dimensional, and where L is a linear bounded transformation on W_s into W. Then we note that in order for $L\omega_s$ to be a random variable it is necessary and sufficient that L be Hilbert-Schmidt. Hence let L be Hilbert-Schmidt. Then $y(\omega_2)$ being Gaussian, it is completely characterized by the corresponding covariance operator:

$$I + LL^*$$

Since LL^* is certainly Hilbert-Schmidt (actually of course trace-class), we can apply the Krein factorization theorem to obtain the representation

$$(I + LL^*)^{-1} = (I - \mathscr{L}^*)(I - \mathscr{L})$$

where $\mathscr{L}$ is a Hilbert-Schmidt Volterra operator:

$$\mathscr{L}f = g; \; g(t) = \int_0^t k(t,s)f(s)ds \quad \text{a.e. } 0 < t < T$$

mapping W into itself. In particular we note that

$$z(\omega_2) = y(\omega_2) - \mathscr{L}y(\cdot)$$

also defines white noise; and defining

$$(I + M) = (I - \mathscr{L})^{-1}$$

where M must then be also Hilbert-Schimdt and Volterra, we note that we can represent $y(\omega_2)$ also as:

$$y(\omega_2) = Mz(\omega_2) + z(\omega_2) \tag{3.9}$$

In this form we can seek the derivative of the weak distributions induced by y(.) to Gauss measure (induced by $z(\omega_2)$ but the processes are no longer independent. However it is shown in [10] that the derivative is a random-variable if and only if

$$M + M^*$$

is trace-class. But, in the present instance this readily follows from the fact that LL^* is trace-class, since

$$LL^* = MM^* + (M + M^*)$$

In other words in the model (3.9), the conditions that M be trace-class is always satisfied if it is deduced from the model [3.1]. Incidentally, it is of interest to note that the derivative is given by:

$$g(\omega) = \text{Exp} - 1/2 \, [||M\omega||^2 - 2\,[M\omega,\omega] + \text{Tr}(M + M^*)] \tag{3.10}$$

and can be deduced from (3.4). Also it should be noted that

$$\text{Tr}(M + M^*) = \text{Tr}\,(\mathscr{L} + \mathscr{L}^*)$$

and also

$$= E[||x - \mathscr{L}y||^2]; \; x = L\omega_s \tag{3.11}$$

The last formula is particularly interesting since it has a variational interpretation. Since $L\omega_s$ is such that the covariance LL^* is trace-class we can formulate the problem of minimizing

$$E[||L\omega_s - Ky(\omega_2)||^2] \tag{3.12}$$

over the class of all Hilbert-Schmidt Volterra operators K. But to show that a minimum exists and is given by the H. S. Volterra operator K_o, it is enough to show that

$$\frac{d}{d\lambda} E(||L\omega_s - (K_o + \lambda K)y(\omega_2)||^2]\Big|_{\lambda = 0} = 0$$

$$= \frac{d}{d\lambda} \sum_1^\infty E[L\omega_s - (K_o + \lambda K)y(\omega_2), \phi_i]^2\Big|_{\lambda = 0} = 0$$

or,

$$\frac{d}{d\lambda} \text{Tr. } (K_o + \lambda K)(I + LL^*)(K_o + \lambda K)^* - 2 LL^*(K_o + \lambda K)^*\} = 0$$

which yields

$$\text{Tr. } (K_o(I + LL^*) - LL^*)K^* = 0$$

for which it is necessary and sufficient that

$$K_o(I + LL^*) - LL^*$$

be the ajoint of a H.S. Volterra operator. But substituting $\mathscr{L}$ for K_o, we see that

$$\begin{aligned} & y(I + LL^*) - LL^* \\ &= (\mathscr{L} - I)(I + LL^*) + I \\ &= -(I - \mathscr{L}^*)^{-1} + I \\ &= -(I + M^*) + I \\ &= M^* \end{aligned}$$

Hence $\mathscr{L}$ yields the optimal minimising H. S. Volterra operator. The main point to be noted here is that existing of an optimal solution to the minimisng problem (3.12) is equivalent to that of the Krein factorization. Whether L is Volterra or not plays no role.

Conditional Expectation: Bayes Formula

Let us note now one important by-product of Theorem 2.1. Let $\phi(\cdot)$ be any element of W. Then by

$$E[[f(\omega_s), \phi] \mid y(\omega_2)] \tag{3.13}$$

we shall mean the limit of the Cauchy sequence (in the mean of order two):

$$E[[F(\omega_s), \phi] \mid P_n y(\omega_2)] \tag{3.14}$$

where P_n is a sequence of monotone increasing finite dimensional projections converging strongly to the Identity. It is implicit that this limit is independent of the particular sequence P_n chosen. We can then state: (Baye's Formula)

Theorem 4.1

$$E[[f(\omega_s), \phi] \mid y(\omega_2)]$$

$$= \frac{\int_W [S,\phi] \; \mathrm{Exp} - 1/2 \; \{||S||^2 - 2 \; [S, y(\omega_2)]\} \; d\mu_S}{\int_W \mathrm{Exp} - 1/2 \; \{||S||^2 - 2 \; [S, y(\omega_2)]\} \; d\mu_S} \qquad (3.15)$$

Remark Note that (3.15) is defined for every y in W.

Proof

Given the monotone sequence of finite dimensional projections $\{P_n\}$, we may consider an orthonormal basis $\{\phi_n\}$ for W such that P_n corresponds to the space spanned by the first n. Then we can calculate

$$E[[f(\omega_s), \phi_1] \mid P_n y(\omega_2)]$$

by the (finite-dimensional) Bayes rule:

$$= \frac{\int_W [S, \phi_1] \; \mathrm{Exp} - 1/2 \; \{||P_n S||^2 - 2 \; [P_n S, y]\}. \; d\mu_S}{\int_W \mathrm{Exp} - 1/2 \; \{||P_n S||^2 - 2 \; [P_n S, y]\} \; d\mu_S}$$

and obtain in the limit, the formula (3.15) with ϕ_1 for ϕ. The formula for arbitrary ϕ is then immediate therefrom.

Corollary

Let P(t) denote the Projections W onto W(t). Then for any ϕ in W, and $0 \leq t \leq T$,

$$E \; [P(t) \; f(\omega_s), P(t)\phi] \mid P(t)y(\omega_2)]$$

$$= \frac{\int_W [P(t)S, P(t)\phi] \; \mathrm{Exp} - 1/2 \; \{||P(t)S||^2 - 2 \; [P(t)S, P(t)y]\}. \; d\mu_S}{\int_W \mathrm{Exp} - 1/2 \; \{||P(t)S||^2 - 2 \; [P(t)S, P(t)y]\}. \; d\mu_S} \qquad (3.16)$$

Proof The proof is immediate

Likelihood Ratio: General Case

Let us now consider the general case where the signal process is not necessarily Gaussian. Let

$$y(t) = S(t) + N(t) \qquad 0 \le t \le T < \infty$$

where $S(\cdot)$ and $N(\cdot)$ are independent processes. We shall assume that the signal $S(\cdot)$ has finite energy: (corresponding to 3.3)

$$\int_0^T E(||S(t)||^2)dt < \infty$$

For each t, $0 < t \le T$, let

$$W(t) = L_2[R_n; (0,t)]$$

We shall shorten W(T) to simply W. Under condition (3.3), the process $S(\cdot)$ induces a countably additive measure on W (and hence on W(t) for each t). [The cylinder measure on W can be extended to be countably additive, in other words; this is a consequence of the Sazonov theorem]. Thus $y(\cdot)$ defines a weak distribution on W defined by the characteristic function:

$$E[e^{i[y,h]}] = C_s(h) \text{ Exp } - 1/2 \; ||h||^2 \tag{3.17}$$

where

$$C_s(h) = E[e^{i[S,h]}]$$

where we have used the inner-product notation:

$$[S,h] = \int_0^T [S(t), h(t)dt, h \in W.$$

Then the cylinder measure μ_y induced by $y(\cdot)$ is absolutely continuous with respect to Gauss measure μ_G and the Radon-Nikodym derivative is defined by the function:

$$f(\omega) = \int_W \text{Exp} - 1/2 \{||S||^2 - 2 [S,\omega)\} d\mu_s \tag{3.18}$$

Thus for any cylinder set C,

$$\mu_y(C) = \underset{n \to \infty}{\text{limit}} \int_C f(P_n\omega)\, d\mu_G$$

where P_n is any sequence of finite dimensional projections strongly convergent to the identity.

Let $\{\phi_n\}$ be an orthonormal basis in W and let L denote the mapping of W into ℓ_2:

$$Lx = a;\ a_n = \int_0^T [x(\sigma),\ \phi_n(\sigma)]d\sigma.$$

Let

$$LS = \zeta$$

Let μ_ζ denote the measure induced on ℓ_2 by this mapping. Then we can rewrite (3.18) in the form

$$f(\omega) = \int_{\ell_2} \text{Exp} - 1/2 \{[\zeta,\zeta] - 2 [\zeta,\ L\omega]\}\, d\mu_\zeta \tag{3.19}$$

It must be emphasised that (2.6) is defined for every element ω in W. Note also that (3.19) can be defined with respect to any orthornormal system $\{\phi_n\}$.

Let us next consider the likelihood functional f(y) where $y(\cdot)$ is the observation. For this purpose, let (3.19) be defined with respect to the orthornormal system $\{\phi_n\}$. For each t, $0 < t \leq T$, define the operators $\Lambda(t)$, mapping into ℓ_2 by:

$$\Lambda(t)x = a;\ a_n = \int_0^t [\phi_n(\sigma),\ x(\sigma)]d\sigma$$

Let

$$R(t) = \Lambda(t) \quad \Lambda(t)^*.$$

Then the Radon-Nikodym derivative of the measure induced by the process $y(\cdot)$ over $[0,t]$ with respect to Gauss measure on $W(t)$ is given by:

$$f(t,\omega) = \int_{\ell_2} \text{Exp} - 1/2 \quad \{[R(t) \zeta,\zeta] - 2 [\zeta, \Lambda(t)\omega]\} \, d\mu_\zeta \tag{3.20}$$

Note that $\Lambda(T) = L$. Let P_n denote the projection operator corresponding to the first n basis functions $\{\phi_i\}$, $i = 1,\ldots n$. Then we define

$$\hat{\zeta}(t) = \lim_n \; E[\zeta | \Lambda(t) \quad P_n y]$$

As we have seen, we have (Bayes Formula) that

$$\hat{\zeta}(t) = \frac{\int_{\ell_2} \zeta \; \text{Exp} - 1/2 \quad \{[R(t)\zeta,\zeta] - 2 [\zeta, \Lambda(t)y]\} \, d\mu_\zeta}{\int_{\ell_2} \text{Exp} - 1/2 \; \{[R(t)\zeta,\zeta] - 2 [\zeta, \Lambda(t)y]\} \, d\mu_\zeta}$$

Note that, by Schwartz Inequality

$$||\hat{\zeta}(t)||^2 \leq \frac{\int_{\ell_2} ||\zeta||^2 \; \text{Exp} - 1/2 \; \{[R(t)\zeta,\zeta] - 2 [\zeta, \Lambda(t)y]\} \, d\mu_\zeta}{\int_{\ell_2} \text{Exp} - 1/2 \quad \{[R(t)\zeta,\zeta] - 2 [\zeta, \Lambda(t)y]\} \, d\mu_\zeta}$$

$$= \frac{\int_{\ell_2} ||\zeta||^2 \; \text{Exp} - 1/2 \; ||R(t)\zeta \; - \Lambda(t)y||^2 \; d\mu_\zeta}{\int_{\ell_2} \text{Exp} - 1/2 \; ||R(t)\zeta - \Lambda(t)y||^2 \; d\mu_\zeta}$$

$$\leq c \, E[||\zeta||^2] \; \text{Exp} + 1/2 \; (||\Lambda(t)y|| + k)^2 \quad 0 < c < \infty \; , \; 0 < k < \infty \tag{3.21}$$

It should be noted that such an estimate is not available in the Wiener process version. Moreover we shall show that (3.20) is actually obsolutely continuous in t with an L_2-derivative. Let $\phi(t)$ be infinitely differentiable with compact support in (0,T). Then

$$\int_0^T [f(t,\omega)\ \phi'(t)]dt$$

$$= \int_{\ell_2} \int_0^T [\text{Exp} - 1/2\ \{[R(t)\zeta,\zeta] - 2\ [\zeta,\ \Lambda(t)\omega]\}\ \phi'(t)dt]\}\ d\mu_\zeta$$

$$= \int_{\ell_2} \left(\int_0^T - 1/2\ ||\sum_1^\infty \phi_i(t)\ \zeta_i||^2 + [\sum_1^\infty \phi_i(t)\ \zeta_i,\ \omega(t)] \right) \left(\text{Exp} - 1/2\ \{[R(t)\zeta,\zeta] - 2\ [\zeta,\ \Lambda(t)\omega]\}\ \phi(t)dt \right) d\mu_\zeta$$

where we note that both

$$||\sum_1^\infty \phi_i(t)\zeta_i||^2 \quad \text{and} \quad [\sum_1^\infty \phi_i(t)\zeta_i,\ \omega(t)]$$

are in L_2 [0,T] for each ζ in ℓ_2. Hence the derivative is (defined a.e. $0 < t < T$):

$$\int_{\ell_2} (- 1/2\ ||\sum_1^\infty \phi_i(t)\zeta_i||^2 + [\sum_1^\infty \phi_i(t)\zeta_i,\ \omega(t)])\ \ \text{Exp} - 1/2\ \{[R(t)\zeta,\zeta] - 2\ [\zeta,\ \Lambda(t)\omega]\}\ d\mu_\zeta$$

we shall next prove that

$$g_N(t) = \sum_1^N \phi_i(t)\hat{\zeta}_i(t) \qquad 0 \le t \le T$$

converges in the norm of W. But this is immediate from the fact that analogous to (3.21):

$$||g_N(t)||^2 \le c\ E[||\sum_1 \phi_i(t)\zeta_i||^2]\ \text{Exp} + 1/2\ ||\Lambda(t)y||^2 \qquad \text{a.e. } 0 < t < T$$

Let

$$\hat{S}(t) = \sum_1^\infty \phi_i(t)\hat{\zeta}_i(t)$$

and

$$\widehat{||S(t)||^2} = \frac{\int_{\ell_2} ||\sum_1^\infty \phi_i(t)\zeta_i||^2 \ \text{Exp} - 1/2 \ \{[R(t)\zeta,\zeta] - 2\ [\zeta, \Lambda(t)y]\} \ d\mu_\zeta}{\int_{\ell_2} \text{Exp} - 1/2 \ \{[R(t)\zeta,\zeta] - 2\ [\zeta, \Lambda(t)y]\} \ d\mu_\zeta}$$

Then from (2.13) we can write:

$$\frac{d}{dt} \text{Log } f(t, y) = -1/2 \ \{||\hat{S}(t)||^2 - [\hat{S}(t), y(t)2 + \widehat{||S(t)||^2} - ||\hat{S}(t)||^2\}$$

and hence finally, for the log likelihood functional:

Log f(y)

$$= -1/2 \left\{ \int_0^T ||\hat{S}(t)||^2 dt - 2 \int_0^T [\hat{S}(t), y(t)]dt \right.$$

$$\left. + \int_0^T [\widehat{||S(t)||^2} - ||\hat{S}(t)||^2]dt \right\} \qquad (3.22)$$

we note that the third term can also be expressed as

$$\lim_{n \to \infty} E[||S(t) - \hat{S}(t)||^2 \ \Big| \Lambda(t)P_n y]$$

The formula (3.22) differs from the Wiener process version in the appearance of the thir term; in the case where S(t) is Gaussian, we know that this reduces to

$$E[||S(t) - \hat{S}(t)||^2]$$

which is then also independent of the observation y(·) as we have already seen.

4. Dynamic Systems.

Finite Dimensional Case:

We wish now to specialise our results to the case where $S(\theta,t)$ has a stochastic differential system representation:

$$S(\theta,t) = C(\theta)\ x(\theta,t)$$

$$\frac{dx(\theta,t)}{dt} = A(\theta)\ x(\theta;t) + F(\theta)\ \omega(t);\ x(\theta,0) = 0. \tag{4.1}$$

and the observation process has the form:

$$y(\theta,t) = S(\theta,t) + G\ \omega(t) \tag{4.2}$$

where we shall first consider the finite dimensional case so that $C(\theta)$, $A(\theta)$, $F(\theta)$, G are all rectangular matrices with,

$$A(\theta):\ m \times m,\quad F(\theta):\ m \times n,$$

$$F(\theta)G^* = 0$$

$$GG^* = \text{Identity matrix}$$

We take $\omega(\cdot)$ as sample functions of white noise in

$$W = L_2[(0,T);\ R_n].$$

Now equation (4.1) for each fixed θ has (see [10]) the unique solution.

$$x(\theta,t) = \int_0^t e^{A(\theta)\ (t-s)}\ F(\theta)\ \omega(s)ds \quad 0 < t < T$$

and

$$x(\theta,t) = L\omega$$

defines a Hilbert-Schmidt operator on W into

$$W_S = L_2((0,T);\ R_m)$$

In that case

$$\hat{S}(\theta,t) = C(\theta)\ \hat{x}\ (\theta,t)$$

where

$$\dot{\hat{x}}(\theta,t) = A(\theta)\ \hat{x}(\theta,t) + P(\theta;t)\ C(\theta)^{*}[y(t) - \hat{x}(\theta,t)]\ \hat{x}(\theta,0) = 0 \quad (4.3a)$$

and $P(\theta;t)$ satisfies the (Riccati) equation:

$$\dot{P}(\theta,t) = A(\theta)\ P(\theta,t) + P(\theta,t)\ A^{*}(\theta)$$

$$+ F(\theta)\ F^{*}(\theta) - P(\theta,t)\ (\theta)^{*}\ C(\theta)\ P(\theta,t);$$

$$P(\theta,0) = 0 \quad (4.3b)$$

And finally, the likelihood functional becomes:

$$\text{Exp} - 1/2 \left\{ \int_0^T ||\hat{S}(\theta,t)||^2 dt - 2 \int_0^t [\hat{S}(\theta,t)\ y(t)]dt \right.$$

$$\left. + \int_0^T \text{Tr. } C(\theta)\ P(\theta,t)\ C(\theta)^{*} dt \right\} \quad (4.4)$$

This result was apparently first obtained by F. C. Schweppe [11] by proceeding formally from the time-discrete case.

In estimating parameters by maximising (4.4) we proceed as follows. Let

$$q(\theta,T) = \int_0^T ||\hat{S}(\theta,t)||^2 dt - 2 \int_0^T [\hat{S}(\theta,t),\ y(t)]dt$$

$$+ \int_0^T \text{Tr. } C(\theta)\ P(\theta,t)\ C(\theta)^{*} dt$$

we use the iteration (modified Newton-Raphson):

$$\theta_{n+1} = \theta_n - M(\theta_n)^{-1} \nabla_\theta q(\theta_n;T)$$

where $M(\theta;T)$ is the matrix with components:

$$m_{ij}(\theta) = \int_0^T [\frac{\partial}{\partial \alpha_i} \hat{S}(\theta,t), \frac{\partial}{\partial \alpha_j} \hat{S}(\theta,t)]dt$$

where $\{\alpha_i\}$ denote the 'components' of θ. We assume that $M(\theta,T)$ is positive definite on the set of admissible parameters θ.

Infinite-Dimensional Case

The extension of (4.1) to the infinite dimensional case (corresponding to partial differential equations) can take many forms. One version is treated in [10]. For each θ, $A(\theta)$ in (4.1) is now the infinitesimal generator of a strongly continuous semigroup over a separable Hilbert space H_S. Equation (4.1) remains formally the same, with $F(\theta)$ being a linear bounded transformation for each θ, mapping H into H_S, $\omega(\cdot)$ denoting white noise in

$$W = L_2[(0,T]; H]$$

H being a separable Hilbert space.

Similarly $C(\theta)$ is assumed to be linear bounded and G linear bounded with

$$F(\theta)G^* = 0; \; GG^* = \text{Identity}$$

In that case the finite dimensional version (4.4) goes over without change provided we assume that

$$\int_0^T E(||C(\theta)\, x\, (\theta;t)||^2)dt < \infty$$

This in particular implies that

$$C(\theta)\ P(\theta,t)\ C(\theta)^*$$

is trace-class a.e. and that

$$\int_0^T \mathrm{Tr.}C(\theta)\ P(\theta,t)\ C(\theta)^* dt < \infty.$$

However from the practical point of view we need to consider the case where $C(\theta)$ is allowed to be unbounded, and uncloseable, (corresponding to 'boundary' or 'pointwise' observations in distributed parameter systems).

Here we shall consider such an extension that takes care of the application to the case of turbulence with non-rational spectrum (see Sec. 6). Actually the model we shall study represents a wide variety of situations assuming only linearity. Thus we take:

$$y(t) = S(t,\theta) + n_1(t) \quad 0 < t < T \tag{4.5}$$

where $n_1(\cdot)$ is white noise in

$$W_0 = L_2((0,T);\ R_m)$$

and $S(t,\theta)$ has the form

$$S(t,\theta) = \int_0^t B(\theta;\ t-s)\ \ u(S)ds + \int_0^t F(\theta;\ t-s)n_2(s)ds \tag{4.6}$$

where $n_2(\cdot)$ is white-noise (independent of $n_1(\cdot)$) in

$$W_S = L_2((0,T);\ R_n),$$

$u(\cdot)$ is a known (deterministic) function, and $\int_0^\infty ||u(t)||^2 dt < \infty$

and for each θ:

$$\int_0^\infty ||B(\theta;\sigma)||^2 d\sigma + \int_0^\infty ||F(\theta;\sigma)||^2 d\sigma < \infty \qquad (4.7)$$

Note that (4.1), (4.2) form a special case of (4.5), (4.6), (4.7), where the Laplace transforms of $B(\theta,\sigma)$ and $F(\theta,\sigma)$ are constrained to be rational functions. To handle the generalization when one (or both) is not necessarily rational, we proceed (see [12]) as follows. We show that we can rewrite (4.7) in terms of a partial differential equation representation. Thus let

$$H = L_2[0, \infty; R_p]$$

where p is the dimension of the observation. Let A denote the generator of the shift-semigroup over H:

$$\mathcal{D}(A) = [f \in H \mid f(\cdot) \text{ is absolutely continuous and the derivative } f'(\cdot) \in H]$$

and

$$Af = f'$$

Let u(t) be an m x 1 matrixfunction. Let $B(\theta)$ denote a linear bounded operator mapping R_m into H defined by:

$$B(\theta)\, u = g;\; g(t)=B(\theta,t)u \quad 0 < t < \infty,\ u \in R_m$$

Let

$$\omega(t) = \begin{pmatrix} n_1(t) \\ n_2(t) \end{pmatrix}$$

so that $\omega(\cdot)$ is white noise in

$$L_2((0,T);\ R_p \times R_n)$$

Let $F(\theta)$ denote the linear bounded operator mapping $R_p \times R_n$ into H defined by

$$F(\theta)v = g;\ g(t) = F(\theta;t)n_2,\ v = \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}, \quad \begin{matrix} n_1 \ \varepsilon\ R_p \\ n_2 \ \varepsilon\ R_n \end{matrix}$$

Finally let G denote the mapping of $R_p \times R_n$ into R_p defined by

$$Gv = w,\ w = n_1,\ v = \begin{matrix} n_1 \\ n_2 \end{matrix},\ n_1\ \varepsilon\ R_p,\quad n_2\ \varepsilon\ R_n.$$

Then we claim that (4.6) is representable as:

$$\begin{aligned} \dot{x}(\theta,t) &= A\,x\,(t) + B(\theta)u(t) + F(\theta)\,\omega(t);\ x(\theta,0) = 0 \\ y(t) &= C\,x(\theta,t) + G\omega(t) \end{aligned} \tag{4.8}$$

where it should be noted that

$$F(\theta)G^* = 0;\ GG^* = I$$

where C is the operator defined by

$$\text{Domain of } C = [f\ \varepsilon\ H \mid f(\cdot) \text{ is continuous in } 0 \leq t < \infty]$$

$$Cf = f(0)$$

We assume that $B(\theta,t)$, $F(\theta,t)$ are locally continuous in $0 \leq t < \infty$. We can readily see that $x(\theta,t)$ is then in the domain of C for each t. That (4.8) is the same as (4.7) follows from the representation:

$$x(\theta,t)=\int_0^t S(t-\sigma)\, B(\theta)u(\sigma)d\sigma + \int_0^t S(t-\sigma)\, F(\theta)\, \omega(\sigma)d\sigma$$

where $S(t)$ is the semigroup (shift) generated by A. Even though C is not closeable, $C\, x(\theta,t)$ is defined and is locally continuous in $0 \leq t < \infty$, for each $u(\cdot)$ and $\omega(\cdot)$. We can then (see [12]) deduce the analogue of (4.3a), (4.3b) as:

$$\dot{\hat{x}}(\theta,t) = A\,\hat{x}(\theta,t) + (C\,P(\theta,t))^* [y(t) - C\,\hat{x}(\theta,t)]\quad \hat{x}(\theta,0) = 0 \tag{4.9}$$

where $P(\theta,t)$ satisfies:

$$[\dot{P}(\theta,t)\, x,\, y] = [P(\theta,t)x,\, A^*y] + [P(\theta,t)y,\, A^*x]$$

$$+ [F(\theta)^*x,\, F(\theta)^*y]$$

$$- [CP(\theta,t)x,\, CP(\theta,t)y]$$

$$P(\theta,0) = 0, \tag{4.10}$$

$x, y \in$ Domain of A^*.

In particular $P(\theta,t)$ maps into the domain of C, and

$$C\,P(\theta,t)$$

is linear bounded (even though C is not closed; see [12]) for each t. Moreover (cf[12]):

$$(C\,P(\theta,t))^* \in \text{Domain of } C$$

and

$C\,(CP(\theta,t))^*$ is bounded (and automatically trace-class being finite dimensional)

The Radon-Nikodym derivative formula (4.4) now becomes

$$\text{Exp} - 1/2 \int_0^T ||\hat{S}(\theta,t)||^2 dt - 2 \int_0^T [\hat{S}(\theta,t), y(t)]dt$$

$$+ \int_0^T \text{Tr.} \;\; C(C\, P(\theta,t))^* dt \qquad (4.11)$$

In this version it is important to note that the 'steady-state' solution of (4.10) exists:

$$P(\theta, \infty)x = \lim_{t\to\infty} P(\theta, \infty)x$$

$$0 = [P(\theta, \infty)x, A^* y] + [P(\theta, x)y, A^* x]$$

$$+ [F(\theta)^* x, F(\theta)^* y] - [C\, P(\theta, \infty)x, C\, P(\theta, \infty)y] \qquad (4.12)$$

provided

$$\int_0^\infty \int_0^\infty ||F(\theta, \sigma+t)||^2 \, d\sigma \;\; dt \; < \infty \qquad (4.13)$$

5. Application

We turn now to an application: estimation of stability and control derivatives from flight test data. The dynamic system considered arises from the longitudinal mode perturbation equations for an aircraft in windgust (turbulence) (Rediess Taylor, see [13]). We use the Dryden version of the spectrum of turbulence, which is rational, so that the total system is finite dimensional. Leaving the many essential details to the comprehensive work of Iliff [11], the state space formulation of the problem is as follows: (see also [12]):

$$\dot{x}(t) = A\, x(t) + B\, u(t) + F\, n_2(t)$$

$$V(t) = C\, x(t) + D\, u(t) + G\, n_1(t)$$

where $n_1(\cdot)$, and $n_2(\cdot)$ are independent whit Gaussian, and the matrices in the equations have the form:

$$A = \begin{bmatrix} Z_1 & 0 & 1 & Z_1 \\ 0 & 0 & 1 & 0 \\ M_1 & 0 & M_3 & M_1 \\ 0 & 0 & 0 & -\frac{\bar{v}}{1000} \end{bmatrix}$$

$$B = \begin{bmatrix} Z_4 \\ 0 \\ M_4 \\ 0 \end{bmatrix} \quad F = \sigma \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{20\bar{v}} \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{10M_1-\bar{v}Z_1}{g} & 0 & 10M_3 & \frac{10M_1-\bar{v}Z_1}{g} \\ k_1 & 0 & \frac{32K_1}{\bar{v}} & k_1 \end{bmatrix}$$

$$D = \begin{bmatrix} 0 \\ 0 \\ \frac{10M_4-\bar{v}Z_4}{g} \\ 0 \end{bmatrix}$$

g = acceleration due to gravity

G = diag. [.0005, .0001, .01, .0001]

The lettered entries (stability and control derivatives) are unknown, except for $\bar{v}$, which is known (=1670). Note that the turbulence power is an unknown parameter also.

The sampling interval was 0.02 seconds while the data bandwidth is about $5H_z$. Figure 1 shows the complete time history of the observation v(t) (four components) subdivided into various regions for later identification, as well as the input time-history. Estimates were computed over the various subregions each by three methods:

Method I: Neglecting the measurement noise on the angle of attack measurment (v_4) and following the corresponding maximal likelihood technique developed in [13]. This is reasonable for this particular example at high turbulence levels.

Method II: This is the method developed herein.

Method III: This was a "check" method, in which the turbulence was ignored completely in the model.

The results are summarized in Figure 2. Sample means and variances of the estimates obtained over the different data-regions are shown, along with the wind-tunnel values as well estimates obtained on other turbulence-free (smooth air) flights. It can be seen that Method II yields the most consistent estimates. It also turns out that Method II is the least in computational time -- the estimates converging in fewer iterations. It can also be seen that ignoring the turbulence leads to the worst results. For more discussion see [11]. The remaining figures indicate the nature of the "fit" obtained using the estimated coefficients to the observed data. Figure 3 shows the close agreement provided by Method II. Figures 4 and 5 indicate how much worse the agreement is on the same stretch of data if the turbulence is not accounted for.

If we use the non-rational (Kolmogorov) version of the spectrum of turbulence, we have to use [4.11]. In particular in this case,

$F(\theta,t)$ has the form

$$F(\theta,t) = (a(\theta)\, t^{5/6} + b(\theta)\, t^{-1/6})\, e^{-kt}$$

corresponding to the spectral density of the form (cf[15]):

$$\frac{1 + c\, f^2}{(1 + d\, f^2)^{11/6}}$$

The possibility of using flight-test data to distinguish between the two models of the spectral density is an intriguing one at the present time.

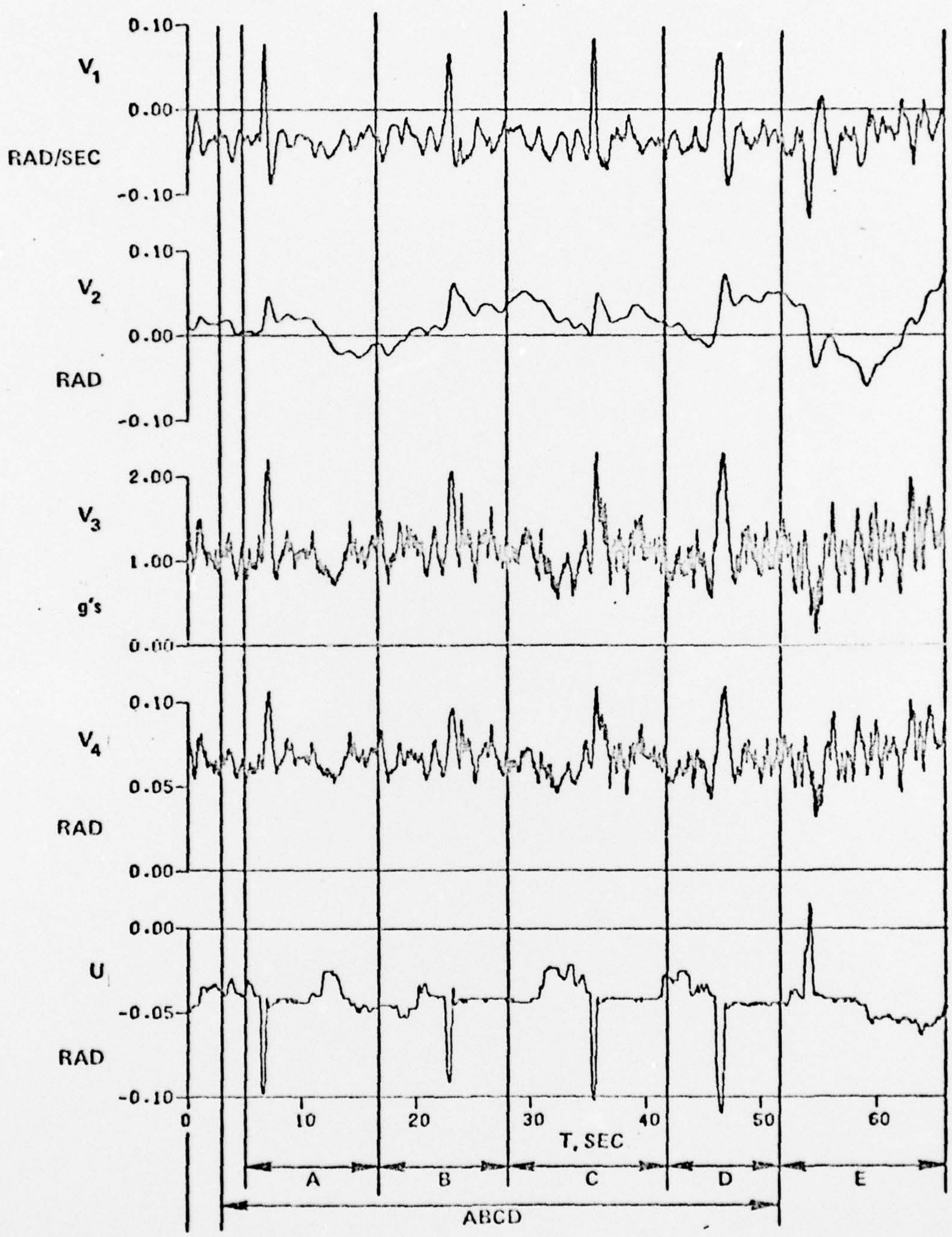

Figure 1. Total Jetstar Turbulence Time History Showing Intervals of Each Maneuver

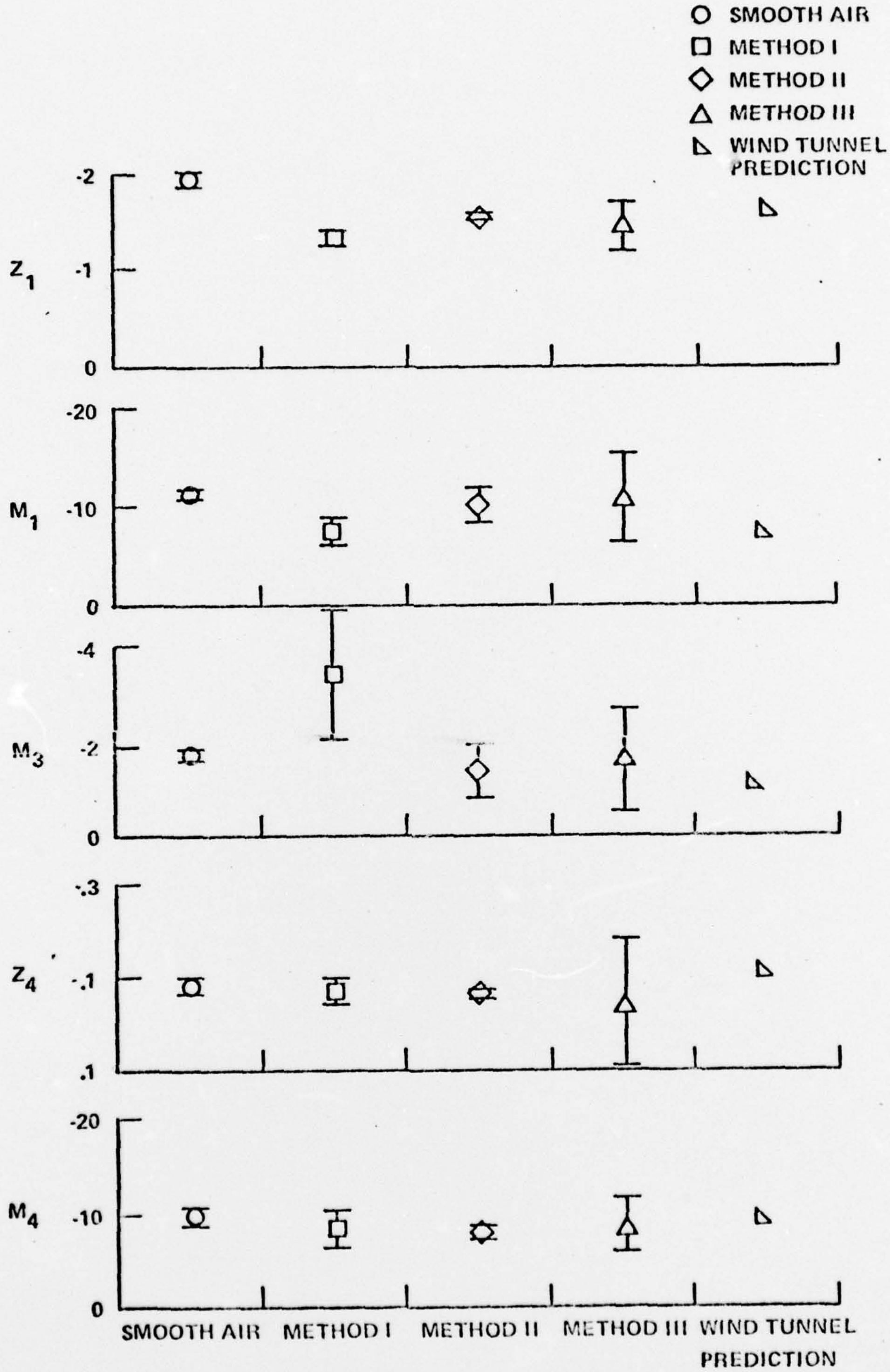

Figure 2. Means and Standard Deviations for Five Methods of Estimating Coefficients

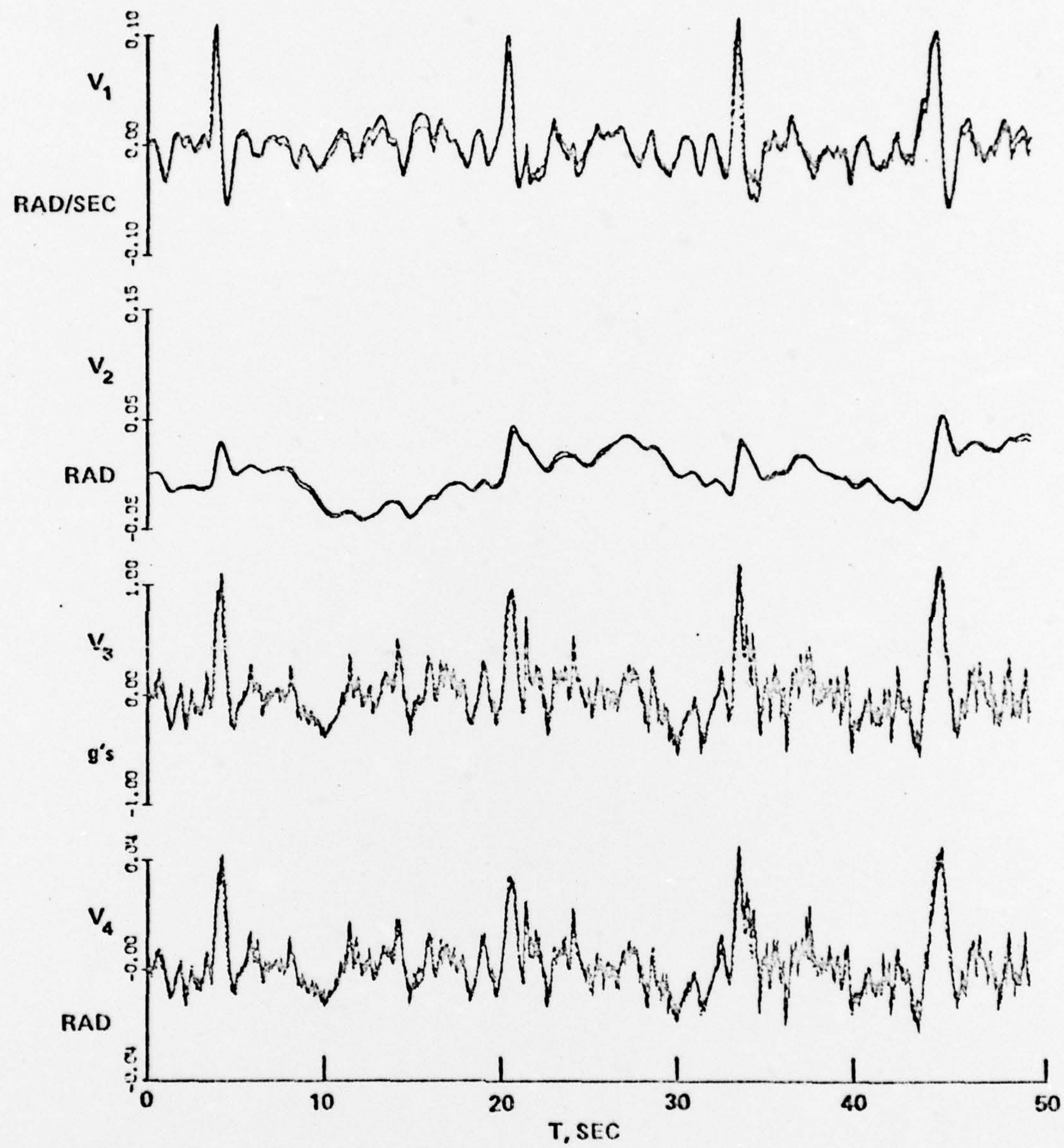

Figure 3. Comparison of Flight Data from Maneuver ABCD and the Estimated Data – Method II

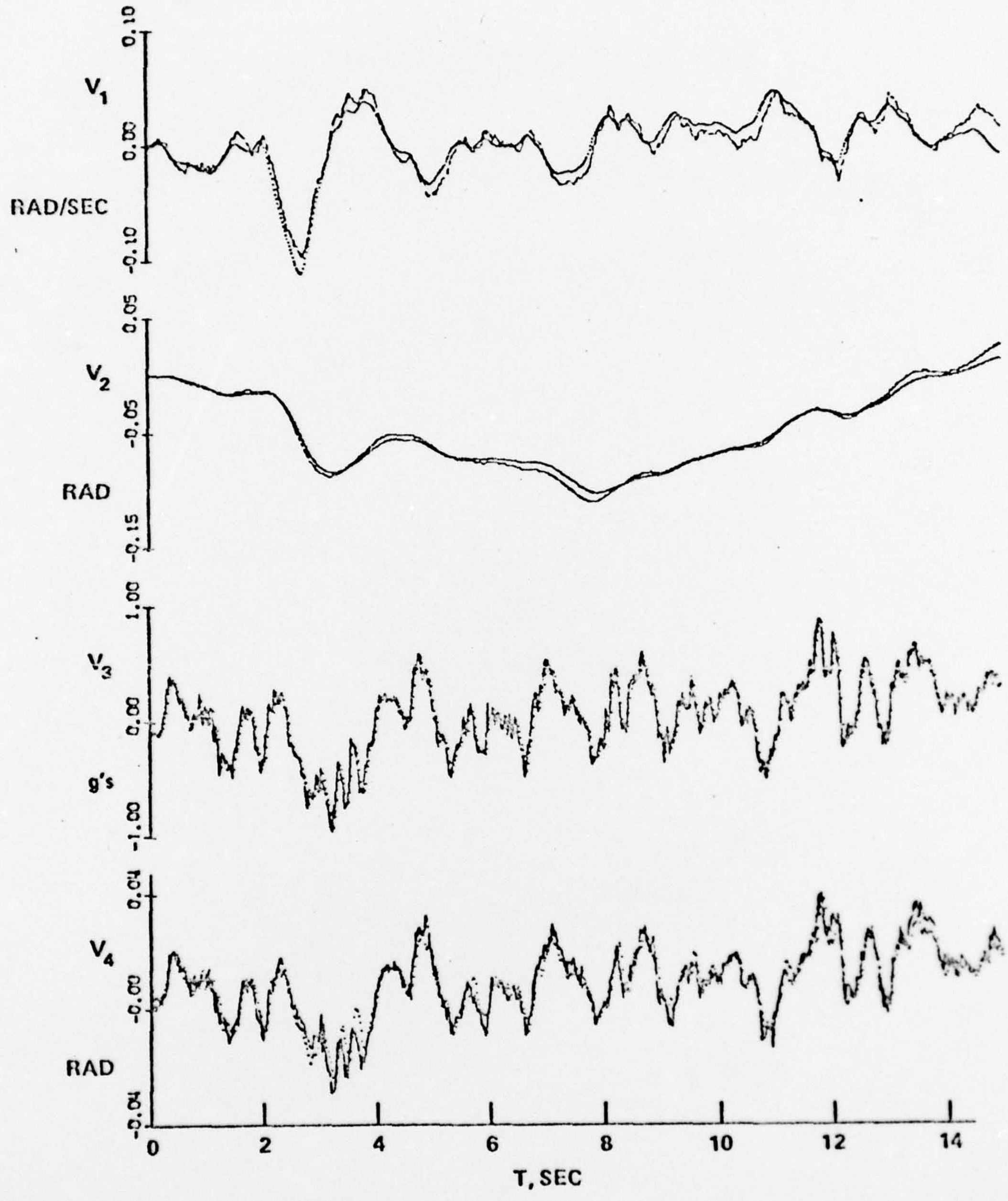

Figure 4. Comparison of Flight Data from Maneuver E and Estimated Data -- Method II

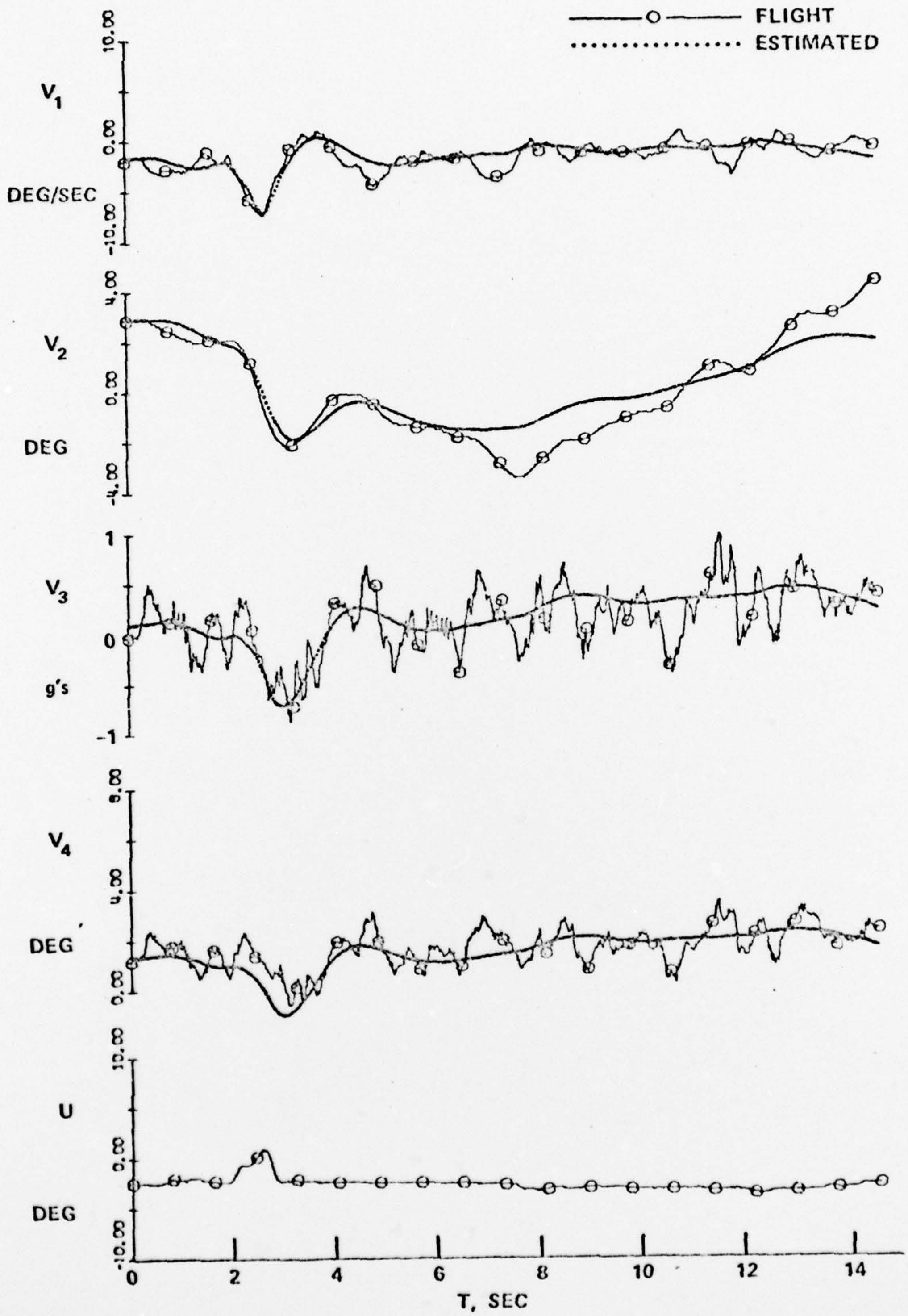

Figure 5. Comparison of Flight Data from Maneuver E and the Estimated Data Obtained by Method III

References

1. Proceeding of the IFAC Symposia on Identification and System Parameter Estimation, 1976, 1973, 1970.
2. Y. Lipster and A. Shiryayev: Statistic of Random Processes, Nanka, 1974.
3. A. V. Skorolehod: Integration in Hilbert Space, Springer-Verlag, 1975.
4. L. Gross: Harmanic Analysis on Hilbert Space, Memoir No. 46, A.M.S., 1963.
5. V. Sazonov: Remarkian Characteristic functionals, Theory of Probability and Applications, Vol. 3, 1958.
6. A. V. Balakrishnan: A White Noise Version of the Girsanov Formula, Proceedings of the Symposium on Stochastic Differential Equations, Kyoto, Japan, July 1976.
7. T. E. Duncan: "Evaluation of likelihood functionals" Information and Control, Vol. 13, No. 1, July 1968.
8. G. Kallianpur and C. Striebel: Stochastic differential equations occurring in the estimation of continuous parameter stochastic processes, Theory of Probability and Applications, Vol. 14, 1969.
9. Sosulin, Yu and Stratanovich, K. C., "Optimum Detection of a Diffusion Process in White Noise", Radio Tech. i. Electronic. Vol. 10 (English Translation Vol. 10, p. 704-714).
10. A. V. Balakrishnan: Applied Functional Analysis, Springer-Verlag, 1976.
11. F. C. Schweppe: Evaluation of Likelihood Functions for Gaussian Signals, IEEE Transactions on Information Theory, Vol. 11, p. 61-70.
12. A. V. Balakrishnan: Stochastic Filtering and Control of Linear Systems, A Geneve Theory, Proceedings of the Conference on Control of Partial Differential Equations, Naval Surface Weapons Center, Maryland, 1976.
13. A. V. Balakrishnan: Identification and Adaptive Control, An Application to Flight Control Systems, Journal of Optimization Theory and Applications, Marhc 1972.

14. K. W. Iliff, Dissertation, School of Engineering UCLA, 1973, (UCLA - ENG - 7340, May 1973).

15. Houbolt, J. C. et al: "Dynamic Response of Airplanes to Atmospheric Turbulence including flight data on input and response", NASA TR R-199, June 1964.